Bayesian Nonparametric Predictive Inference and Bootstrap Techniques

By

Pietro Muliere and Piercesare Secchi
Università degli Studi di Pavia and University of Minnesota

Technical Report #597
February, 1994

1 Introduction

The Bayesian bootstrap was introduced by Rubin [1981] as a convenient Bayesian analogue of the bootstrap technique created by Efron [1979]. Our first aim will be to show that Rubin's bootstrap and Efron's bootstrap are first order equivalent from the predictive point of view. First order asymptotic equivalence between the two bootstrap procedures was proved by Lo [1987] in the sense that for almost all sample sequences they achieve the same limiting conditional distribution.

We then investigate the question as to whether the posterior distributions obtained by means of the bootsrap procedures arise via Bayes Theorem from a prior on the class of distribution functions. The fact that the Bayesian bootstrap "gives zero probability to the event that a future observation is unequal to the observed values in the sample" led some Bayesian authors to question its applicability and to suggest modifications to the basic procedure [Meeden1993]. We also suggest a new generalization of the Bayesian bootstrap which takes into account prior opinions and has moreover the appealing property that the predictive distribution for a future observation is not necessarily concentrated on the observed values.

The paper is organized as follows. After few preliminaries and a section devoted to two characterizations of the Dirichlet process, Efron's and Rubin's bootstraps will be introduced and examined from the predictive perspective in the fourth section. In section 5 we tackle the problem regarding the existence of a prior on the space of distribution functions consistent with the posterior distributions produced by the bootstrap techniques. A new Bayesian resampling plan is proposed in section 6. A couple of applications in the last section concludes the paper.

2 Preliminaries

Let Beta(α, β) indicate a beta distribution function with density f defined, for any $x \in \Re$, by

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I[0 < x < 1],$$

where $\alpha > 0$ and $\beta > 0$. If $\alpha = 0$ and $\beta > 0$, then Beta$(0, \beta)$ is defined to be the distribution function of the point mass at 0; if $\alpha > 0$ and $\beta = 0$, then Beta$(\alpha, 0)$ is defined to be the distribution function of the point mass at 1.

The rest of this section collects, without proof, few well known facts concerning the beta distribution which we will need in the sequel.

If α and β are two non negative quantities such that at least one is different from 0, the r-th moment of a Beta(α, β) distribution is:

$$m_r(\alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)}, & \alpha > 0 \\ 0, & \alpha = 0 \end{cases}$$

for $r = 1, 2, \ldots$.

Let d be a distance on the space of distribution functions such that $\lim_{n\to\infty} d(G_n, G) = 0$ if and only if G_n converges in distribution to G as n grows to infinity. Indicate with $N(\mu, \sigma^2)$ the Normal distribution with mean μ and variance σ^2.

2.1 **Lemma**. *For any n, let $G_n = Beta(\alpha_n, \beta_n)$ where $\alpha_n > 0$ and $\beta_n > 0$. If*

$$\lim_{n\to\infty} \frac{\alpha_n}{\alpha_n + \beta_n} = \gamma \quad \text{and} \quad \lim_{n\to\infty} \frac{n}{\alpha_n + \beta_n} = l < \infty,$$

then

$$\lim_{n\to\infty} d(\, G_n \,, \, N(\frac{\alpha_n}{\alpha_n + \beta_n}, \frac{l\gamma[1-\gamma]}{n})\,) = 0.$$

2.2 **Lemma**. *For any n, let $G_n = Beta(\alpha_n, \beta_n)$. Assume that*

$$\lim_{n\to\infty} \alpha_n = \alpha, \quad \lim_{n\to\infty} \beta_n = \beta$$

where α and β are two non negative quantities such that at least one is different from zero. Then

$$\lim_{n\to\infty} d(\, G_n \,, \, \text{Beta}(\alpha, \beta)\,) = 0.$$

3 Dirichlet process prior: two characterizations

Let $\{X_n\}$ be an exchangeable sequence of real random variables (r.v.) defined on a probability space $(\Omega, \mathcal{F}, P)$. De Finetti's Representation Theorem guarantees the existence of a random distribution function F conditionally on which the variables of the sequence $\{X_n\}$ are independent and identically distributed (i.i.d.) with distribution F. An interesting prior distribution for F was introduced by Ferguson [1973] in a fundamental paper on a Bayesian approach to nonparametric statistics. We will indicate this prior, called Dirichlet process, by $\mathcal{D}(\alpha)$ where α is a monotone increasing, right continuous, real valued function defined on $\Re$ such that $\alpha(-\infty) = 0$ and $\alpha(+\infty) = k > 0$. By representing α as the product of the constant $k > 0$ and a proper distribution function F_0, we can interpret F_0 as the prior guess at F and k as the 'measure of faith' in this guess. For the definition of the Dirichlet process and a rewiew of its salient features we refer to Ferguson [1973, 1974]. In the sequel we will make use of the following characterizations of this process.

Let F be a random distribution function and, for any Borel set B, indicate with $F(B)$ the probability assigned to B by F.

3.1 **Characterization (Doksum 1973).** *If for any n and any Borel set B, the posterior distribution of $F(B)$, given a sample $X_1, \ldots, X_n$ from F, depends on the sample only through the number of observations that fall in B, then F is either a Dirichlet process or of one of the following types:*

T_1 : *F is degenerate at a given distribution function ($F \equiv F_0$);*

T_2 : *F is the distribution function of a r.v. concentrated on a random point ($F = I_{[X,\infty)}$, where X has distribution F_0);*

T_3 : *F is the distribution function of a r.v. concentrated on two non-random points ($F = UI_{[a,\infty)} + (1-U)I_{[b,\infty)}$, where U has an arbitrary distribution on $[0,1]$ and $a < b$).*

3.2 **Remark.** Processes of types T_1 and T_2 can be viewed as limit of a Dirichlet process $\mathcal{D}(kF_0)$ as $k \to \infty$ and $k \to 0$ respectively. $\diamond$

The next characterization requires an assumption analogous to the one appearing in Characterization 3.1 restricted now to posterior expectations. Assume that, for any $n \geq 1$, $X_1, \ldots, X_n$ is a sample from the distribution F. Define F_n to be the empirical distribution function of $X_1, \ldots, X_n$ by setting, for any $x \in \Re$,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[X_i,\infty)}(x).$$

3.3 **Characterization (Regazzini 1978, Lo 1991).** *Let F_0 be a distribution function and $k > 0$. Then F is a Dirichlet process $\mathcal{D}(kF_0)$ if and only if, for any $n \geq 1$ and any Borel set B,*

$$P[\, X_{n+1} \in B \,|\, X_1, \ldots, X_n \,] = \frac{k}{k+n} F_0(B) + \frac{n}{k+n} F_n(B).$$

Finally and for the sake of completness, we recall the following conjugate property of the Dirichlet process.

3.4 **Proposition (Ferguson 1973).** *Let F be a Dirichlet process $\mathcal{D}(kF_0)$ and $X_1, \ldots, X_n$ a sample from F. Then the posterior distribution of F, given $X_1, \ldots, X_n$, is the Dirichlet process $\mathcal{D}(kF_0 + nF_n)$.*

4 Predictive inference and the bootstrap

From the completely predictive point of view most inferential problems regarding an exchangeable sequence of r.v.'s $\{X_n\}$ reduce to the computation of the conditional probability

$$P[X_{n+1} \in B \mid X_1, \ldots, X_n]$$

where B is a Borel set. However, the exchangeability assumption implies that

$$P[X_{n+1} \in B \mid X_1, \ldots, X_n] = E[F(B) \mid X_1, \ldots, X_n] \tag{4.1}$$

where F is the random distribution function conditionally on which the r.v.'s of the sequence $\{X_n\}$ are i.i.d..

The bootstrap procedures provide methods for approximating the conditional distribution

$$\mathcal{L}(\phi(F, \mathbf{X}) \mid X_1, \ldots, X_n)$$

where, for clarity of exposition, we indicated with $\mathbf{X}$ the sample $X_1, \ldots, X_n$ and $\phi(F, \mathbf{X})$ is a functional depending on both F and $\mathbf{X}$.

In this section we want to discuss the question as to when it is reasonable to approximate the expected value appearing in equation (4.1) by means of a bootstrap procedure applied to the conditional distribution $\mathcal{L}(F(B) \mid X_1, \ldots, X_n)$. We will consider both the original bootstrap procedure introduced by Efron [1979] and the Bayesian bootstrap procedure proposed by Rubin [1981].

Efron's bootstrap. Given a sample $X_1, \ldots, X_n$ from F the approximation suggested by Efron's bootstrap is the following:

$$\mathcal{L}(\phi(F, \mathbf{X}) \mid X_1, \ldots, X_n) \approx \mathcal{L}(\phi(F_n^*, \mathbf{X}^*) \mid X_1, \ldots, X_n) \tag{4.2}$$

where $\mathbf{X}^*$ indicates a sample $X_1^*, \ldots, X_n^*$ from the empirical distribution function F_n of $X_1, \ldots, X_n$ and F_n^* is the empirical distribution function of the sample $\mathbf{X}^*$. In particular, for any Borel set B,

$$\mathcal{L}(F(B) \mid X_1, \ldots, X_n) \approx \mathcal{L}(F_n^*(B) \mid X_1, \ldots, X_n).$$

It's easy to verify that, for any Borel set B,

$$\mathcal{L}(nF_n^*(B) \mid X_1, \ldots, X_n) = \text{Binomial}(n, F_n(B)). \tag{4.3}$$

Therefore, by means of Efron's bootstrap, one obtains

$$P[X_{n+1} \in B \mid X_1, \ldots, X_n] \approx F_n(B). \tag{4.4}$$

Rubin's bootstrap. Let $\{V_n\}$ be a sequence of i.i.d. random variables with exponential distribution of parameter 1 and assume that $\{V_n\}$ is independent of $\{X_n\}$. The Bayesian bootstrap procedure proposed by Rubin suggests the following approximation:

$$\mathcal{L}(\phi(F, \mathbf{X}) \mid X_1, \ldots, X_n) \approx \mathcal{L}(\phi(F_n^R, \mathbf{X}) \mid X_1, \ldots, X_n)$$

where, given $X_1, \ldots, X_n$, F_n^R is the random distribution function defined, for any $x \in \Re$, by setting

$$F_n^R(x) = \frac{1}{\sum_{i=1}^n V_i} \sum_{i=1}^n V_i I_{[X_i,\infty)}(x).$$

In particular,

$$\mathcal{L}(\, F(B) \,|\, X_1, \ldots, X_n \,) \approx \mathcal{L}(\, F_n^R(B) \,|\, X_1, \ldots, X_n \,).$$

However, it is immediately verified that, for any Borel set B,

$$\mathcal{L}(\, F_n^R(B) \,|\, X_1, \ldots, X_n \,) = \mathrm{Beta}(nF_n(B), n[1 - F_n(B)]). \tag{4.5}$$

Therefore, by use of the Bayesian bootstrap, one obtains again the approximation

$$P[\, X_{n+1} \in B \,|\, X_1, \ldots, X_n \,] \approx F_n(B).$$

Both bootstrap techniques propose to approximate the probability that a future observation X_{n+1} falls in B, given the past observations $X_1, \ldots, X_n$, with the frequency according to which the past observations fall in B. In this sense Efron's bootstrap and the Bayesian bootstrap of Rubin are first order equivalent from the predictive point of view.

4.6 **Remark.** First order asymptotic equivalence of Efron's and Rubin's bootstraps has been already shown by Lo [1987]. In fact, one can show that, for any Borel set B,

$$\lim_{n\to\infty} d(\, \mathcal{L}(\, F_n^*(B) \,|\, X_1, \ldots, X_n \,)\,,\, N(\, F_n(B)\,,\, \frac{F(B)[1 - F(B)]}{n}\,)\,) = 0$$

on a set of probability one. However, because of Lemma 2.1, the same result is true when $\mathcal{L}(\, F_n^*(B) \,|\, X_1, \ldots, X_n \,)$ is replaced by $\mathcal{L}(\, F_n^R(B) \,|\, X_1, \ldots, X_n \,)$. $\diamond$

5 Bootstrap and prior distribution for F

Given a random distribution function F, let $\{X_n\}$ be a sequence of random variables conditionally i.i.d. with distribution F. The Bayesian approach to the computation of

$$P[\, X_{n+1} \in B \,|\, X_1, \ldots, X_n \,]$$

requires to elicit a prior distribution for F on the space of distribution functions and then use the posterior distribution of F to compute the expected value appearing in (4.1). It is thus important to investigate the question as to when the approximations proposed by the bootstrap procedures are in agreement with this approach. This will be the aim of this section.

5.1 **Lemma.** *Let $k > 0$ and F_0 be a proper distribution function. Then F is a Dirichlet process $\mathcal{D}(kF_0)$ if and only if, for any n and for any Borel set B,*

$$\begin{aligned} \mathcal{L}(\,F(B)\,|\,X_1,\ldots,X_n\,) &= \\ &= \mathrm{Beta}(kF_0(B)+nF_n(B), k[1-F_0(B)]+n[1-F_n(B)]) \end{aligned} \tag{5.2}$$

where F_n is the empirical distribution function of $X_1,\ldots,X_n$.

Proof. Necessity follows from the definition of the Dirichlet process and Proposition 3.4 while sufficiency is a direct consequence of Characterization 3.3. ◇

By means of Lemma 2.1 again, one can show that, on a set of probability one, the distance d between the distribution function (5.2) and $N(F_n(B), n^{-1}F(B)[1-F(B)])$ goes to zero when n grows to infinity. Therefore the asymptotic results of Remark 4.6 show that is reasonable to approximate the posterior distribution of $F(B)$ by means of a bootstrap procedure when n is large and the prior distribution of F is a Dirichlet process. Note that, when n is large, the weight given to the prior opinion, elicited in the parameter F_0 of the posterior Dirichlet process, becomes negligible.

For n fixed and $k \to 0$, the posterior distribution function (5.2) converges according to d to a $\mathrm{Beta}(nF_n(B), n[1-F_n(B)])$ which is the approximation of $\mathcal{L}(\,F(B)\,|\,X_1,\ldots,X_n\,)$ proposed by Rubin's bootstrap. This shows that the Bayesian approach and Rubin's procedure are in agreement when F is a Dirichlet process and the weight k given to the prior opinion F_0 is extremely small. However Dirichlet process is not defined for $k = 0$. The following theorem investigates the question as to when a prior distribution for F exists such that (4.5) is a true posterior distribution for $F(B)$.

5.3 **Theorem.** *Given a random distribution function F, let $\{X_n\}$ be a sequence of r.v.'s conditionally i.i.d. with distribution F. For any Borel set B and for any n,*

$$\mathcal{L}(\,F(B)\,|X_1,\ldots,X_n\,) = \mathrm{Beta}(nF_n(B), n[1-F_n(B)]) \tag{5.4}$$

if and only if F is the distribution function of a r.v. concentrated on a random point.

Proof. With reference to Characterization 3.1, we need to show that (5.4) holds for any B and for any n if and only if F is of type T_2; that is if and only if there is a distribution function G_0 such that

$$F = I_{[X,\infty)} \tag{5.5}$$

where X is a r.v. with distribution function G_0.

Necessity. If (5.4) holds for any Borel set and any n, then, because of Characterization 3.1, F is either a Dirichlet process or of types T_1, T_2 or T_3.

F cannot be a Dirichlet process $\mathcal{D}(kF_0)$, where $k > 0$ and F_0 is a proper distribution function, because of Lemma 5.1.

Assume that F is of type T_1. Then $F \equiv F_0$ where F_0 is a proper distribution function and, for any B and for any n,

$$\mathcal{L}(F(B) \mid X_1, \ldots, X_n) = \mathcal{L}(F_0(B)).$$

But then (5.4) implies that, for any B and for any n,

$$\mathcal{L}(F_0(B)) = \text{Beta}(nF_n(B), n[1 - F_n(B)])$$

and this is possible only if, for any B, $F_0(B) \in \{0, 1\}$ or, equivalently, if $F_0 = F = I_{[a,\infty)}$ where $a \in \Re$.

Finally assume that F is of type T_3. Then, there is a distribution function F_0 on $[0, 1]$ and a r.v. U with distribution F_0 such that

$$F = UI_{[a,\infty)} + (1 - U)I_{[b,\infty)}$$

where $a < b$. Let B be a Borel set such that $a \in B$ and $b \notin B$. Then, for any n,

$$\mathcal{L}(F(B) \mid X_1, \ldots, X_n) = \mathcal{L}(U \mid X_1, \ldots, X_n)$$

and thus, by (5.4),

$$\mathcal{L}(U \mid X_1, \ldots, X_n) = \text{Beta}(nF_n(B), n[1 - F_n(B)]).$$

In particular, for $n = 1$,

$$\mathcal{L}(U \mid X_1) = \begin{cases} I_{[1,\infty)} & \text{if } X_1 = a \\ I_{[0,\infty)} & \text{if } X_1 = b \end{cases}$$

This however is possible only if $P(U \in (0, 1)) = 0$ or, equivalently, if

$$\mathcal{L}(U) = pI_{[0,\infty)} + (1 - p)I_{[1,\infty)}$$

where $0 \leq p \leq 1$. Therefore $F = I_{[X,\infty)}$ where X has distribution function

$$G_0 = (1 - p)I_{[a,\infty)} + pI_{[b,\infty)}.$$

Sufficiency. Let G_0 be a proper distribution function and assume that

$$F = I_{[X,\infty)}$$

where X has distribution G_0. Then, for any n,

$$P[X = X_1 = \ldots = X_n] = 1.$$

Fix a Borel set B and note that, for any n,

$$\mathcal{L}(F(B) \mid X_1, \ldots, X_n) = \begin{cases} I_{[1,\infty)} & \text{if } X_1 \in B \\ I_{[0,\infty)} & \text{if } X_1 \notin B \end{cases}.$$

Therefore, for any n and any Borel set B,

$$\mathcal{L}(F(B) \mid X_1, \ldots, X_n) = \text{Beta}(nF_n(B), n[1 - F_n(B)]).$$

◇

5.6 **Remark.** An analogous result holds for Efron's bootstrap. That is, for any Borel set B and any n,

$$\mathcal{L}(nF(B) \mid X_1, \ldots, X_n) = \text{Binomial}(n, F_n(B))$$

if and only if F is the distribution function of a r.v. concentrated on a random point. The proof of this fact is completely similar to the one given above. ◇

In conclusion: the approximations of $\mathcal{L}(F(B) \mid X_1, \ldots, X_n)$ provided by the bootstrap techniques are 'true' posteriors if the random distribution function F is a process of type T_2. This process however has no interest by itself for a statistician since, for example, it implies that all the variables X_n's are equal with probability one. Nevertheless processes of type T_2 can be considered as limits of Dirichlet processes $\mathcal{D}(kF_0)$ when $k \to 0$. Therefore it seems that, following a Bayesian approach, bootstrap approximations are justifiable either when $k \to 0$ or when the sample size n grows to infinity. In both cases the weight given to the prior opinion becomes negligible. In the next section we want to suggest a new bootstrap technique which has the advantage of taking into account the prior opinion.

6 A new bootstrap technique

In this section we assume that the random distribution function F conditionally on which the r.v.'s of the sequence $\{X_n\}$ are i.i.d. is a Dirichlet process $\mathcal{D}(kF_0)$, wher $k > 0$ and F_0 is a proper distribution function. We want to suggest a resampling procedure with the aim of approximating the conditional distribution

$$\mathcal{L}(\phi(F) \mid X_1, \ldots, X_n) \tag{6.1}$$

where $\phi(F)$ is a functional depending on F. In particular we will consider two types of functional:

Q_1: $\phi_q(F) = \inf\{t \in \Re : F(t) \geq q\}$ where $0 < q < 1$;

Q_2: $\phi_h(F) = \int h \, dF$.

In principle the conditional distribution function (6.1) can be computed by means of Bayes Theorem. For functionals of type Q_2 this has been studied by Hannum, Hollander and Langberg [1981] and by Cifarelli and Regazzini [1990]. Their results, however, are not easy to handle analytically. Therefore arises the need for an approximating technique.

Our proposal depends on the result of the following lemma.

6.2 **Lemma.** *Let F_0 be a discrete distribution function with support $\{z_1, \ldots, z_r\}$ in $\Re$. For $i = 1, \ldots, r$, let p_i be the probability which F_0 assigns to z_i. Assume that $V_1, \ldots, V_r$ are r independent r.v's such that, for $i = 1, \ldots, r$,*

$$\mathcal{L}(V_i) = \text{Gamma}[kp_i, 1]$$

where $k > 0$. Let F be the random distribution function defined, for any $x \in \Re$, by setting

$$F(x) = \frac{1}{\sum_{i=1}^{r} V_i} \sum_{i=1}^{r} V_i I_{[z_i,\infty)}(x).$$

Then F is a Dirichlet process $\mathcal{D}(kF_0)$.

Proof. For any measurable partition $\{B_1, \ldots, B_s\}$ of $\Re$

$$\mathcal{L}(\,F(B_1), \ldots, F(B_s)\,) = \text{Dirichlet}(\,kF_0(B_1), \ldots, kF_0(B_s)\,).$$

Then F is a Dirichlet process $\mathcal{D}(kF_0)$ [Ferguson1973]. $\diamond$

Recall that, by Proposition 3.4, if F is a process $\mathcal{D}(kF_0)$ and $X_1, \ldots, X_n$ is a sample from F, the posterior distribution of F is again a Dirichlet process with parameter $kF_0 + nF_n$. If F_0 is a discrete distribution with finite support, then $(k+n)^{-1}(kF_0 + nF_n)$ is also discrete; let $\{z_1, \ldots, z_r\}$ be the finite support of this last distribution with corresponding probability masses $\{p_1, \ldots, p_r\}$. Then

$$\mathcal{L}(\,\phi(F)\,|\,X_1, \ldots, X_n\,) = \mathcal{L}(\,\phi(\frac{1}{\sum_{i=1}^{r} V_i} \sum_{i=1}^{r} V_i I_{[z_i,\infty)})\,|\,X_1, \ldots, X_n\,)$$

where, given $X_1, \ldots, X_n$, the r.v.'s $V_1, \ldots, V_r$ are independent and such that, for $i = 1, \ldots, r$,

$$\mathcal{L}(V_i | X_1, \ldots, X_n) = \text{Gamma}[(k+n)p_i, 1].$$

In this case it is immediately evident how to apply a Monte Carlo method in order to find an approximation of (6.1). However, in most situations of statistical interest, F_0 is not discrete so that the direct approach just described will not be applicable. When this happens, a possible way out is first to approximate the parameter $kF_0 + nF_n$ with a suitable bounded, monotone increasing, right continuous step function α^* such that $\alpha^*(-\infty) = 0$, and then to use the process $\mathcal{D}(\alpha^*)$ as an approximation of the posterior process $\mathcal{D}(kF_0 + nF_n)$. The Bayesian bootstrap originates from the same idea by setting $\alpha^* = nF_n$. This is equivalent to assigning a negligible weight to the prior guess at F and, as we have seen, it does not cause much harm if the sample size n is large. Our alternative proposal is to approximate $kF_0 + nF_n$ by $(n+k)F^*$ where F^* is the empirical distribution function generated by a bootstrap sample from $(n+k)^{-1}(kF_0 + nF_n)$.

The aim of the rest of this section will be to show that the proposal described above seems reasonable for the functionals ϕ's which satisfy the next condition. Recall that d is a distance on the space of distribution functions which induces the convergence in distribution.

6.3 **Condition.** *For any sequence $\{\beta_n\}$ of distribution functions converging to a distribution β according to d and any $k > 0$,*

$$\lim_{n\to\infty} d(\,\mathcal{L}(\phi(G_n))\,,\,\mathcal{L}(\phi(G))\,) = 0$$

if G_n and G are two Dirichlet processes $\mathcal{D}(k\beta_n)$ and $\mathcal{D}(k\beta)$ respectively.

The following two theorems specify under what assumptions functionals of type Q_1 and Q_2 satisfy Condition 6.3.

6.4 **Theorem.** *Let $0 < q < 1$, $k > 0$ and assume that $\{\beta_n\}$ is a sequence of distribution functions converging to a distribution β according to the distance d. Then*

$$\lim_{n\to\infty} d(\,\mathcal{L}(\phi_q(G_n))\,,\,\mathcal{L}(\phi_q(G))\,) = 0$$

if G_n and G are two Dirichlet processes $\mathcal{D}(k\beta_n)$ and $\mathcal{D}(k\beta)$ respectively.

Proof. Note that for any $t \in \Re$

$$P(\,\phi_q(G_n) \le t\,) = P(\,G_n(t) \ge q\,).$$

We need two consider two cases.

First assume that t is a continuity point for β. Since $\mathcal{L}(\,G_n(t)\,) = \text{Beta}(\,k\beta_n(t)\,,\,k[1 - \beta_n(t)]\,)$ and $\lim_{n\to\infty} \beta_n(t) = \beta(t)$, then, by Lemma 2.2,.

$$\begin{aligned}
\lim_{n\to\infty} P(\,\phi_q(G_n) \le t\,) &= \lim_{n\to\infty} P(\,G_n(t) \ge q\,) \\
&= P(\,G(t) \ge q\,) \\
&= P(\,\phi_q(G) \le t).
\end{aligned}$$

In the second case assume that t is a discontinuity point for β. Then we claim that t is also a discontinuity point for the distribution function of $\phi_q(G)$ so that the theorem is proved. To prove the claim notice that

$$\lim_{\delta\to 0^+} \beta(t-\delta) = \beta(t)^- < \beta(t).$$

Being, for any $\delta > 0$, $\mathcal{L}(\,G(t-\delta)\,) = \text{Beta}(\,k\beta(t-\delta)\,,\,k[1-\beta(t-\delta)]\,)$, by applying Lemma 2.2 again, we obtain

$$\lim_{\delta\to 0^+} P(\,\phi_q(G) \le t - \delta\,) = P(Y \ge q)$$

where Y is a random variable with distribution $\text{Beta}(\,k\beta(t)^-, k[1-\beta(t)^-]\,)$. But $\beta(t)^- < \beta(t)$ so that

$$P(Y \ge q) < P(\,G(t) \ge q\,) = P(\,\phi_q(G) \le t\,).$$

Therefore

$$\lim_{\delta\to 0^+} P(\,\phi_q(G) \le t - \delta\,) < P(\,\phi_q(G) \le t\,)$$

and this proves that t is a discontinuity point for the distribution of $\phi_q(G)$. $\diamond$

6.5 Theorem. *Let $k > 0$ and assume that $\{\beta_n\}$ is a sequence of distribution functions converging to a distribution β according to the distance d. Let h be a real valued, bounded and β-continuous function defined on $\Re$. Then*

$$\lim_{n\to\infty} d(\,\mathcal{L}(\phi_h(G_n))\,,\,\mathcal{L}(\phi_h(G))\,) = 0$$

if G_n and G are two Dirichlet processes $\mathcal{D}(k\beta_n)$ and $\mathcal{D}(k\beta)$ respectively.

Proof. See Corollary 2.7 of Hannum, Hollander and Langberg [1981].

We are now ready for the result which will justify the proposal of the new bootsrap technique described at the beginning of this section. Assume, as before, that, conditionally on a random distribution function F, $\{X_n\}$ is a sequence of r.v.'s i.i.d. with distribution F. Assume also that F is a Dirichlet process $\mathcal{D}(kF_0)$ where $k > 0$ and F_0 is a proper distribution function. Given a sample $X_1, \ldots, X_n$ from F, let $X_1^*, \ldots, X_m^*$ be m random variables conditionally i.i.d. with distribution $(n+k)^{-1}(kF_0 + nF_n)$ where F_n is, as before, the empirical distribution function of $X_1, \ldots, X_n$. Set F_m^* to be the empirical distribution function of $X_1^*, \ldots, X_m^*$ and G_m^* to be a Dirichlet process $\mathcal{D}(\,(n+k)F_m^*\,)$.

6.6 Theorem. *Assume that ϕ is a functional satisfying Condition 6.3. Then,*

$$\lim_{m\to\infty} d(\,\mathcal{L}(\,\phi(G_m^*)\,|\,X_1,\ldots,X_n\,), \mathcal{L}(\,\phi(F)\,|\,X_1,\ldots,X_n\,)\,) = 0$$

on a set of probability one.

Proof. It's enough to notice that, given $X_1, \ldots, X_n$, by Glivenko-Cantelli Lemma,

$$\lim_{m\to\infty} d(\,F_m^*\,,\,(n+k)^{-1}[kF_0 + nF_n]\,) = 0.$$

on a set of probability one. $\diamond$

When m is large, it seems thus reasonable to approximate the conditional distribution $\mathcal{L}(\,\phi(F))\,|\,X_1,\ldots,X_n\,)$ with the conditional distribution $\mathcal{L}(\,\phi(G_m^*)\,|\,X_1,\ldots,X_n\,)$ which in turns can be approximated by means of a Monte Carlo method following the plan described in Lemma 6.2. Details of a resampling procedure supported by this argument will be introduced in the next section along with a couple of numerical examples. Note that the bootstrap technique suggested above approximates the conditional probability $P[\,X_{n+1} \in B\,|\,X_1,\ldots,X_n\,]$ by means of

$$E[\,G_m^*(B)\,|\,X_1,\ldots,X_n\,] = \frac{kF_0(B) + nF_n(B)}{k+n}.$$

This is the same predictive probability obtained by means of Bayes Theorem when F is a Dirichlet process $\mathcal{D}(kF_0)$. Moreover, by taking into account the prior opinion elicited by the distribution F_0 with weight $k > 0$, this resampling plan does not force the future observation to be equal to one of the observed values as was the case with Efron's and Rubin's bootstraps. The technique has also the advantage of being fully consistent with the Bayesian approach since it can be considered as only a tool for approximating numerically a 'true' posterior distribution when this is analytically intractable.

7 Numerical illustrations

We now want to describe a resampling plan which has the aim of computing an approximation for (6.1) and is supported by the arguments of the previous section. The procedure will be tested with two applications.

Assume that a sample $X_1 = x_1, \ldots, X_n = x_n$ has been observed from a random distribution F; for example, de Finetti's Representation Theorem implies that this assumption is correct when $X_1, \ldots, X_n$ are the first n random variables of an exchangeable infinite sequence. Elicit the prior opinion about F by a proper distribution function F_0, the prior 'guess' at F, and by a positive number k, the 'measure of faith' in this guess. In order to build a distribution function which approximates $\mathcal{L}(\,\phi(F)\,|\,X_1 = x_1, \ldots, X_n = x_n\,)$ we will follow these steps:

1. Determine the empirical distribution function F_m^* of m observations $x_1^*, \ldots, x_m^*$ generated by $(n+k)^{-1}(kF_0 + nF_n)$. In particular locate the support $\{z_1^*, \ldots, z_r^*\}$ of F_m^* with corresponding masses $\{p_1^*, \ldots, p_r^*\}$.

2. For $i = 1, \ldots, r$, generate v_i from the distribution function
$$\mathrm{Gamma}((n+k)p_i^*, 1).$$

3. Compute the quantity
$$t = \phi\Big(\frac{1}{\sum_{i=1}^{r}} \sum_{i=1}^{r} v_i I_{[z_i^*,\infty)}\Big).$$

4. Repeat steps (1),(2),(3) s times obtaining the quantities $t_1, \ldots, t_s$.

5. Approximate the conditional distribution function $\mathcal{L}(\,\phi(F)\,|\,X_1 = x_1, \ldots, X_n = x_n\,)$ by means of the empirical distribution function generated by $t_1, \ldots, t_s$.

7.1 **Example.** We observed $x_1 = 0.1, x_2 = 0.05$. Assuming that this is a sample from a random distribution F, we want to compute

$$\mathcal{L}\Big(\int x\, dF(x)\,|\,x_1, x_2\,\Big). \tag{7.2}$$

Our prior guess F_0 at F is a Uniform distribution on $[0, 1]$. To this guess we assigned weights $k = 0, 1, 2, 100$. The procedure described above was then applied with $m = 300$ and $s = 5000$. For different values of k, the distributions approximating (7.2) are summarized in Table 1 by their mean, median, 75th and 95th quantile here indicated with q_{75} and q_{95} respectively.

	Mean			Median		
$k = 0$	[0.0750]	0.0747	(0.0749)	[0.0750]	0.0745	(0.0750)
$k = 1$	[0.2166]	0.2156	(0.2152)	$\star$	0.1789	(0.1786)
$k = 2$	[0.2875]	0.2888	(0.2862)	$\star$	0.2675	(0.2654)
$k = 100$	[0.4916]	0.4918	(0.4891)	$\star$	0.4918	(0.4893)
	q_{75}			q_{95}		
$k = 0$	[0.0875]	0.0871	(0.0872)	[0.0975]	0.0975	(0.0973)
$k = 1$	$\star$	0.2813	(0.2815)	$\star$	0.4770	(0.4804)
$k = 2$	$\star$	0.3688	(0.3667)	$\star$	0.5462	(0.5279)
$k = 100$	$\star$	0.5156	(0.5114)	$\star$	0.5451	(0.5431)

Table 1: Results of the experiments described in Example 7.1

For $k = 0$ our procedure is equivalent to Rubin's bootstrap. However, if the posterior distribution of F, given $X_1 = x_1, X_2 = x_2$, is a Dirichlet process $\mathcal{D}(2F_2)$, then one can verify that (7.2) is a Uniform distribution on $[0.05, 0.1]$ so that the values for the mean, the median and the quantiles can be computed analytically. On the other hand, when the prior distribution of F is a Dirichlet process $\mathcal{D}(kF_0)$, it's always possible to compute the mean of (7.2) [Ferguson1973]. All these analytical results are reported in Table 1 between square brackets.

For comparison purposes we approximated the distribution (7.2) by means of a different technique. Assuming that the prior distribution for F is a Dirichlet process $\mathcal{D}(kF_0)$ and following the markovian procedure suggested by Korwar and Hollander [1973] a sample of size 300 was generated by a Dirichlet process of parameter kF_0+2F_2. The mean of the sample so generated can be viewed as a realization of a random variable having distribution (7.2). The procedure was then iterated 5000 times and the empirical distribution of the sample means thus obtained was considered as an approximation of (7.2). Results relative to this simulation are reported in Table 1 between round brackets.

The results computed by means of these different techniques look all very similar and they all confirm the obvious fact that the more k increases, the more relevant becomes the prior opinion elicited with F_0. $\diamond$

7.3 **Example.** For the purpose of comparing our resampling plan with other bootstrap techniques recently introduced in the Bayesian literature we repeated an experiment originally due to Meeden [1993]. Given a sample $X_1, \ldots, X_n$ from a random distribution function F, we want to estimate the 25th, the 50th and the 75th quantile of F. In the following experiments the sample $X_1, \ldots, X_n$ was generated by a Gamma[20,1]. For $q = 0.25, 0.5, 0.75$, define, as before, the functional

$$\phi_q(F) = \inf\{t \in \Re : F(t) \geq q\}.$$

The exact values of $\phi_q(F)$ when $F = \text{Gamma}[20, 1]$ are reported in Table 2.

Two experiments, differing by the sample size, were performed; in the first one the

	$q=0.25$	$q=0.50$	$q=0.75$
$F=\text{Gamma}[20,1]$	16.83	19.67	22.81

Table 2: Exact values for the quantiles of a Gamma[20,1]

		Prior Distribution			
		$U[0,60]$	$U[8.5,30]$	$\text{LogNormal}[2.97,0.22]$	$\text{Gamma}[20,1]$
		$(n=11,n=25)$	$(n=11,n=25)$	$(n=11,n=25)$	$(n=11,n=25)$
$\bar{q}_{25}$	$k=0$	$(17.21,16.96)$	$(17.21,16.96)$	$(17.21,16.96)$	$(17.21,16.96)$
	$k=1$	$(17.19,16.93)$	$(17.04,16.82)$	$(16.99,16.74)$	$(17.45,16.97)$
	$k=5$	$(16.96,16.87)$	$(16.44,16.42)$	$(16.90,16.81)$	$(17.07,16.92)$
$\bar{q}_{50}$	$k=0$	$(19.89,19.74)$	$(19.89,19.74)$	$(19.89,19.74)$	$(19.89,19.74)$
	$k=1$	$(20.10,19.78)$	$(19.77,19.58)$	$(19.73,19.48)$	$(19.96,19.80)$
	$k=5$	$(20.84,20.21)$	$(19.68,19.42)$	$(19.63,19.55)$	$(19.76,19.68)$
$\bar{q}_{75}$	$k=0$	$(22.68,22.83)$	$(22.68,22.83)$	$(22.68,22.83)$	$(22.68,22.83)$
	$k=1$	$(23.46,23.00)$	$(22.81,22.71)$	$(22.85,22.71)$	$(22.77,22.94)$
	$k=5$	$(27.14,24.21)$	$(23.17,22.74)$	$(22.58,22.69)$	$(22.72,22.87)$

Table 3: Results of the experiments described in Example 7.3

sample size n was set equal to 11, in the second it was set to be 25. We first obtained an approximation of

$$\mathcal{L}(\,\phi_q(F)\,|\,X_1,\ldots,X_n\,) \tag{7.4}$$

by applying the procedure described at the beginning of the section with $m = 100$ and $s = 300$. We then estimated the 25th, the 50th and the 75th quantile of F by the mean of the corresponding distribution (7.4). As prior guess at F we considered four different distributions: Uniform[0,60], Uniform[8.5,30], LogNormal[2.97,0.22] and Gamma[20,1]. To each distribution we assigned three weights: $k = 0, 1, 5$. Each experiment was repeated 100 times; the average values of the estimators, indicated respectively by $\bar{q}_{25}$, $\bar{q}_{50}$ and $\bar{q}_{75}$, are reported in Table 3.

Note again that for $k = 0$ our procedure is equivalent to Rubin's Bayesian bootstrap. Since the values relative to this case were already computed by Meeden, for comparison purposes we reported them in Table 3.

With this experiment Meeden [1993] compared Rubin's bootsrap with the smooth Bayesian bootstrap of Banks [1988] and his own Bayesian bootstrap based on a grid reaching the conclusion that all these procedures performed similarly. Here is a short description of Meeden's technique. Suppose that the random distribution F has support contained in some given interval $I = (c, d)$. Fix a grid $(g_1, \ldots, g_k)$ on I such that $c = g_1 < \cdots < g_k = d$. Given $X_1, \ldots, X_n$, let $B_1, \ldots, B_s$ be the subintervals of the grid which contain at least one observation and, for $j = 1, \ldots, s$, let w_j be the number of observations that fell in subinterval

B_j. Now approximate the posterior distribution of F by the law of a random distribution G whose support is contained in $\bigcup_{j=1}^{s} B_j$. Assume also that G is such that

$$\mathcal{L}(\,G(B_1),\ldots,G(B_s)\,) = \text{Dirichlet}(w_1,\ldots,w_s)$$

and, for $j=1,\ldots,s$, $G(B_j)$ is uniformly spread over B_j.

From the predictive point of view, Meeden's resampling technique has the advantage of not forcing the future observation to be equal to one of the observed values. Notice also that prior opinions are elicited by fixing a grid before the sample is chosen. However Meeden claims that his procedure is noninformative since "the choice of the grid does not seem to matter very much as long as there are no large subintervals which contain lots of probability." Without renouncing to an informative approach, we followed the same criteria when choosing our four prior guesses at F. Table 3 shows that the results obtained with our resampling technique are quite similar to those of Meeden especially when the weight given to the prior opinion is small compared to the sample size. In fact, even for $k=5$ the average values of the estimators are close to those found with $k=0$ with the possible exception of those relative to 75th quantile when the prior is Uniform[0,60]. $\diamond$

8 Acknowledgments

This work was completed while both authors were visiting the School of Statistics and the Institute for Mathemathics and its Applications at the University of Minnesota. The research of Pietro Muliere was supported by C.N.R. AI93.00382.10.

References

BANKS, D.L. (1988), "Histopline smoothing the Bayesian bootstrap", *Biometrika*, 75, 673–684.

CIFARELLI, D.M. and E. REGAZZINI (1990), "Distribution functions of means of a Dirichlet process", *The Annals of Statistics*, 18(1), 429–442.

DOKSUM, K. (1974), "Tailfree and neutral random probabilities and their posterior distributions", *The Annals of Probability*, 2(2), 183–201.

EFRON, B. (1979), "Boostrap methods: another look at the jacknife", *The Annals of Statistics*, 7(1), 1–26.

FERGUSON, T.S. (1973), "A Bayesian analysis of some nonparametric problems", *The Annals of Statistics*, 1(2), 209–230.

FERGUSON, T.S.(1974), "Prior distributions on spaces of probability measures", *The Annals of Statistics*, 2(4), 615–629.

HANNUM, R.C., HOLLANDER, M. and N.A. LANGBERG (1981), "Distributional results for random functionals of a Dirichlet process", *The Annals of Probability*, 9(4), 665–670.

KORWAR, R.M. and HOLLANDER, M. (1973), "Contributions to the theory of Dirichlet processes", *The Annals of Probability*, 1(4), 705–711.

LO, A.Y. (1987), "A large sample study of the Bayesian bootstrap", *The Annals of Statistics*, 15(1), 360–375.

LO, A.Y. (1991), "A characterization of the Dirichlet process", *Statistics and Probability Letters*, 12(3), 185–187.

MEEDEN, G. (1993), "Noninformative nonparametric Bayesian estimation of quantiles", *Statistics and Probability Letters*, 16, 103–109.

REGAZZINI, E. (1978), "Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità", *Giornale dell'Istituto Italiano degli Attuari*, 77–89.

RUBIN, D.B. (1981), "The Bayesian bootstrap", *The Annals of Statistics*, 9(1), 130–134.